

# Dal dato al sapere: knowledge graph e servizi digitali per le scienze del patrimonio nell'esperienza di [DIGILAB@E-RIHS.it](mailto:DIGILAB@E-RIHS.it)

Alberto Bucciero

# Data vs research results.

“ A publication is the **tip** of the iceberg. The **data** is everything underneath. ”

– and yet, only the tip gets archived.

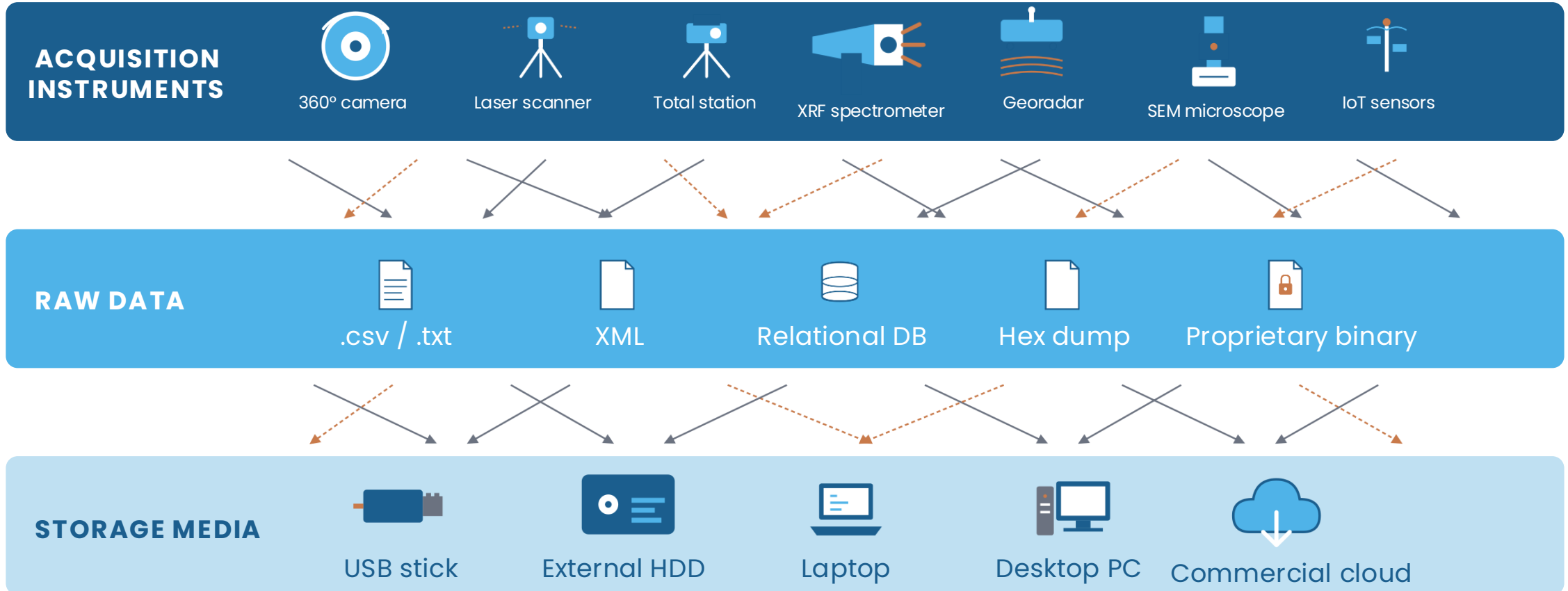
## THREE STRUCTURAL QUESTIONS

- 01 Where are the data used to write an article?
- 02 How were they used?
- 03 How are they related to each other?



PUBLICATIONS AND DATA

# Heritage data is born scattered.



*From fieldwork to publication, 2 to 3 years pass on average. Without a structured system, data rarely survive the journey.*





to knowledge

through information

From data

**preservation**  
in Heritage Science

# Where data physically live shapes who can read them.



## LEGACY MEDIA

Floppy disks, optical media, formats nobody can read anymore.



## PORTABLE DRIVES

USB sticks, SD cards. Travel with a person, vanish with a desk drawer.



## PERSONAL LAPTOP

The researcher's machine. Tied to one person, one project, one career.



## DEPARTMENT NAS

Local servers. Accessible only inside the institution's network.



## SOMEWHERE ELSE

In ten years, will any of these still answer your call?

*Five physical locations, five rules of access – and no guarantee that any survives the project.*

# Commercial clouds: convenient, until they aren't.



**Dropbox**

US-based · consumer-grade



**iCloud**

Apple ecosystem lock-in



**Google Drive**

Free quota, monetised metadata



**OneDrive**

Tied to your Office licence



**AWS**

Pay-per-use forever



**Oracle Cloud**

Enterprise vendor lock-in

**But what about  
LONG-LASTING  
preservation?**

Their business model isn't built for 50 years.  
It's built for the next quarterly report.

*Heritage data outlives every commercial cloud that promises to store it.*

# When a single policy change locks out an entire region.



Official 2025 inaugural portrait of Donald Trump – Daniel Torok / White House.

## noyb (2025): the EU–US Data Privacy Framework is already cracking.

- 01** Democratic members of the PCLOB removed by executive order.
- 02** EU oversight of US data transfers no longer effective.
- 03** EU institutions could lose legal cover for US clouds overnight.

### US Cloud soon illegal? Trump punches first hole in EU-US Data Deal

[Data Transfers](#) / 23 January 2025

Since the Snowden disclosures we know that the US engages in mass surveillance of EU users by scooping up personal data from US Big Tech. The "Privacy and Civil Liberties Oversight Board" (PCLOB) is the key US oversight authority for these laws. [US media now reports](#), that Democratic members of the PCLOB got removed and their email accounts shut down. This brings the number of appointed Members below the threshold to have the PCLOB operate. The fact that the US President simply removed people from an (allegedly) independent authority, question the independence of all other executive redress bodies in the US.

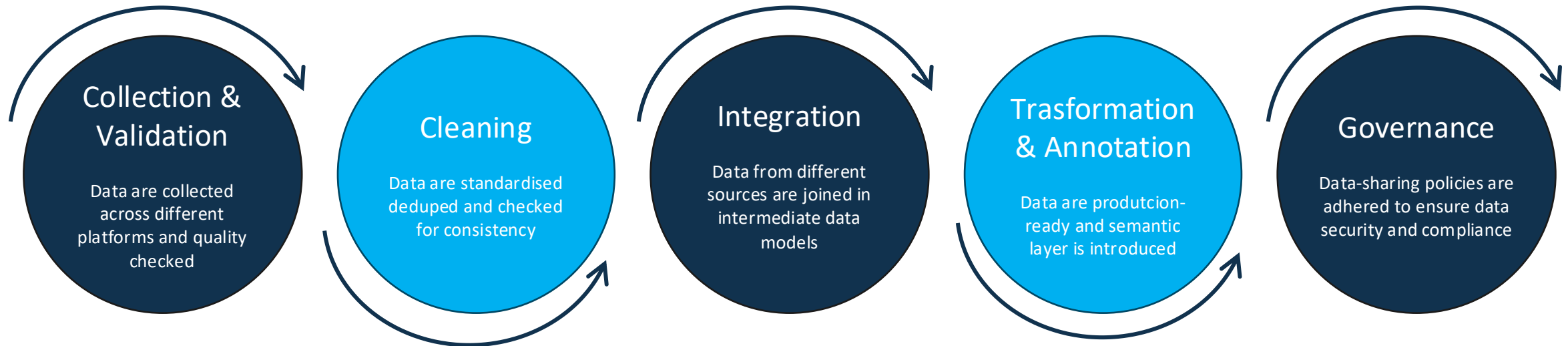
The European Union has relied on these US boards and tribunals to find that the US provides "adequate" protection of personal data. Relying on the PCLOB and other mechanisms, the EU Commission allows European personal data to flow freely to the US in the so-called "Transatlantic Data Privacy Framework" (TADPF). The PCLOB is the only relevant "oversight" element of the deal. Other elements are just acting as recourse bodies. Thousands of EU businesses, government agencies or schools rely on these provisions. Without the TADPF, they would need to stop using US cloud providers like Apple, Google, Microsoft or Amazon instantly.

source: noyb.eu · [us-cloud-soon-illegal-trump-punches-first-hole-eu-us-data-deal](#)

*Cultural heritage data hosted on US commercial clouds is one administrative decision away from being unreachable.*

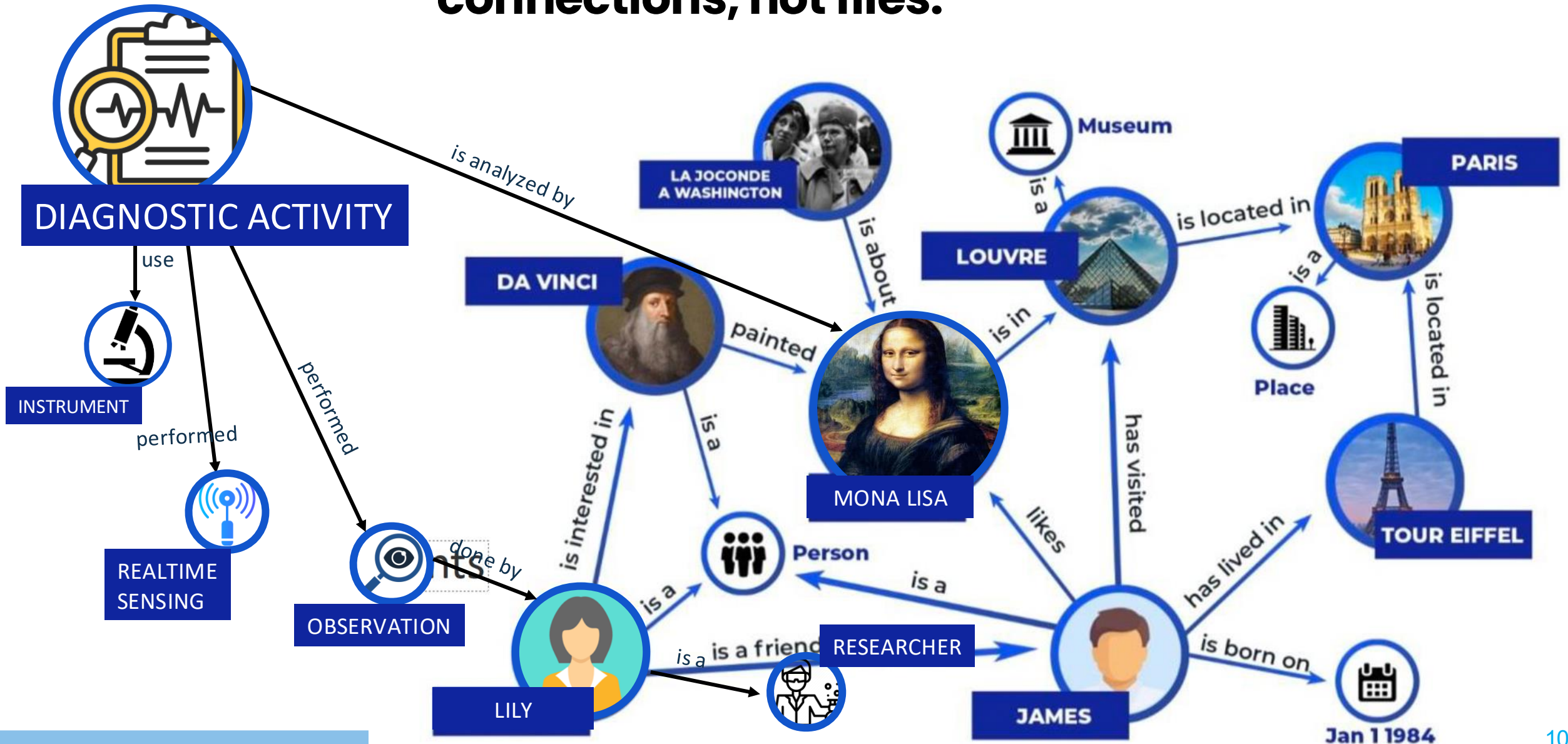
# Curation: data become information only when shaped.

*Five steps that turn raw data into something a peer can actually reuse.*



*Without curation, data stay technically present but practically unreadable.*

# Knowledge requires connections, not files.



# Back to the starting QUESTIONS:

- 01 Where are the data used to write an article?**
- 02 How were they used?**
- 03 How are they related to each other?**



# Q#1: Where are the **data** used to find scientific results presented in publications?



OneDrive



GDrive



Dropbox



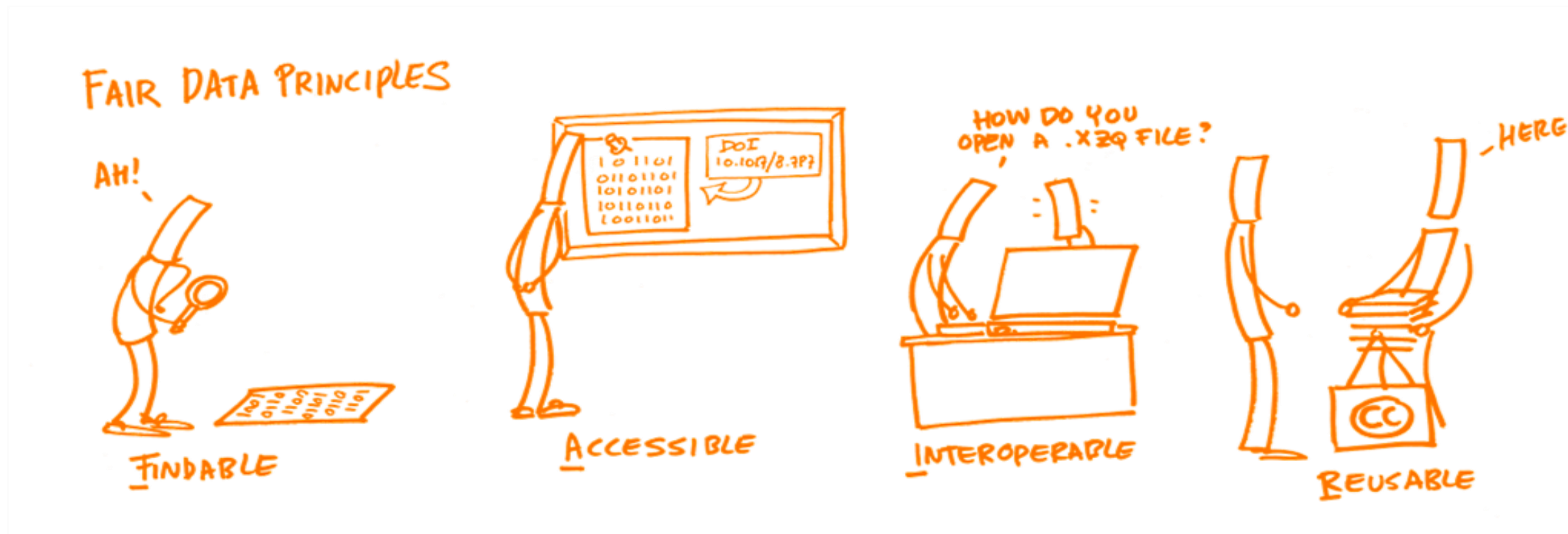
iCloud

Do we really need another VIRTUAL DRIVE?

Is the capability of storing and retrieving data enough ?

# R#1: Preserve DATA FAIRNESS

Is it needed to store all datasets in one or more federated repositories and ensure compliance with the principles of



# scientific data preservation:

# data infrastructures



E-RIHS Cloud Data Center – Lecce- Catania-Napoli



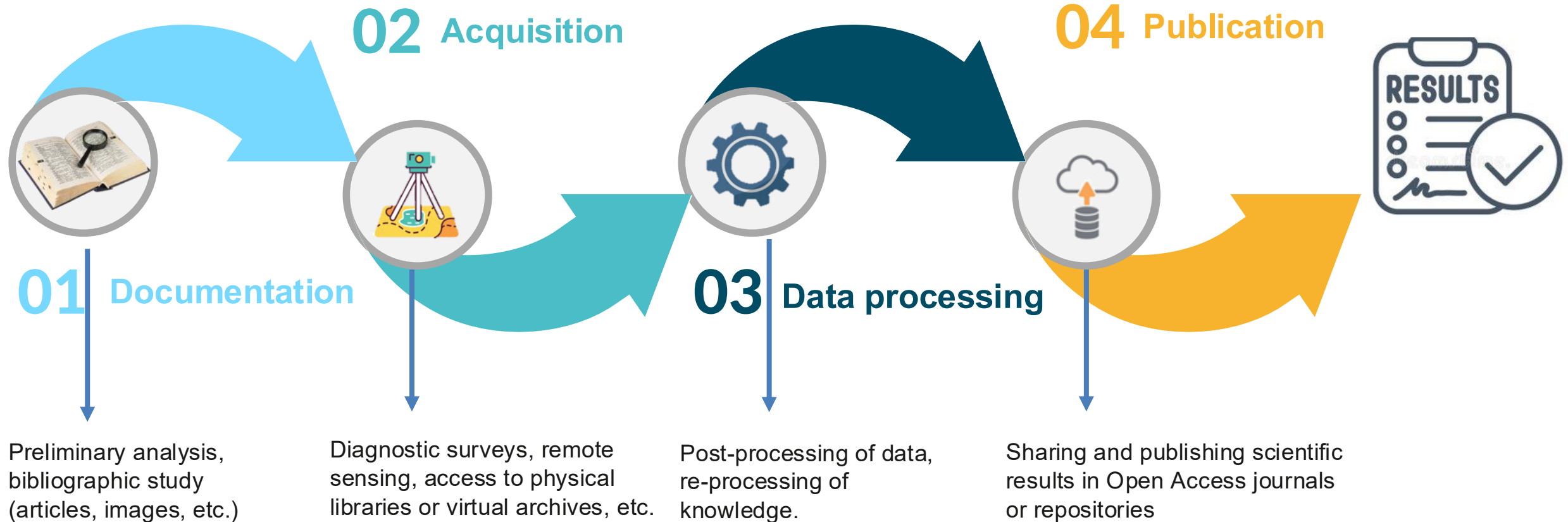
H2IOSC national FEDERATED Data Centers



DARIAH Data Center – Lecce-Catania-Napoli-Roma-Firenze-Pisa

# Q#2: How were **data** used?

## R#2: Save the scientific research **WORKFLOW**



# Q#3: How data they related to each other?

## R#3: Preserve and Discover KNOWLEDGE

The screenshot shows the Dataspace interface with a search bar containing "Observation". The results list includes three entries for "Fourier Transform Mid-infrared (FT-MIR) Spectroscopy of ms. C2792, c.34v (Observation)". A central network graph displays various nodes and relationships. The right sidebar shows "Resource Information" for "Ms. C2792, c. 34v (Physical Thing)" and a list of relationships including "Ms. C2792 (Physical Thing)", "Fourier Transform Mid-infrared (FT-MIR) Spectroscopy of ms. C2792, c.34v (Observation)", "X-ray Fluorescence (XRF) of ms. C2792, c.34v (Observation)", "Imaging of ms. C2792, c.34v (Observation)", and "Video Microscopy of ms. C2792, c.34v (Observation)".

A collage of related data visualizations. It includes a manuscript page with a large illuminated initial 'M', a line graph showing a spectrum, a smaller manuscript page with a similar illuminated initial, and another line graph showing a spectrum. The visualizations are arranged in a layered, overlapping manner.

Dataspace <https://dataspace.ispc.cnr.it>



# E-RIHS

EUROPEAN RESEARCH INFRASTRUCTURE  
FOR HERITAGE SCIENCE

## Access to research facilities, archives and data

E-RIHS provides access to facilities, techniques and archives distributed throughout Europe and organized into four integrated physical and virtual platforms: **ARCHLAB**, **MOLAB**, **FIXLAB** and the new entry **DIGILAB-IT**, implemented in **H2IOSC** for the Italian node of E-RIHS

### E-RIHS PLATFORMS



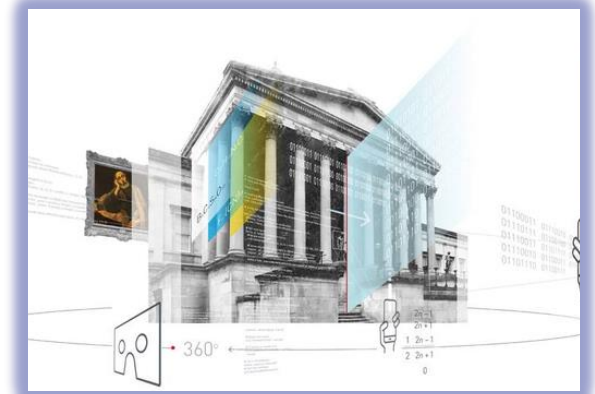
**ARCHLAB**



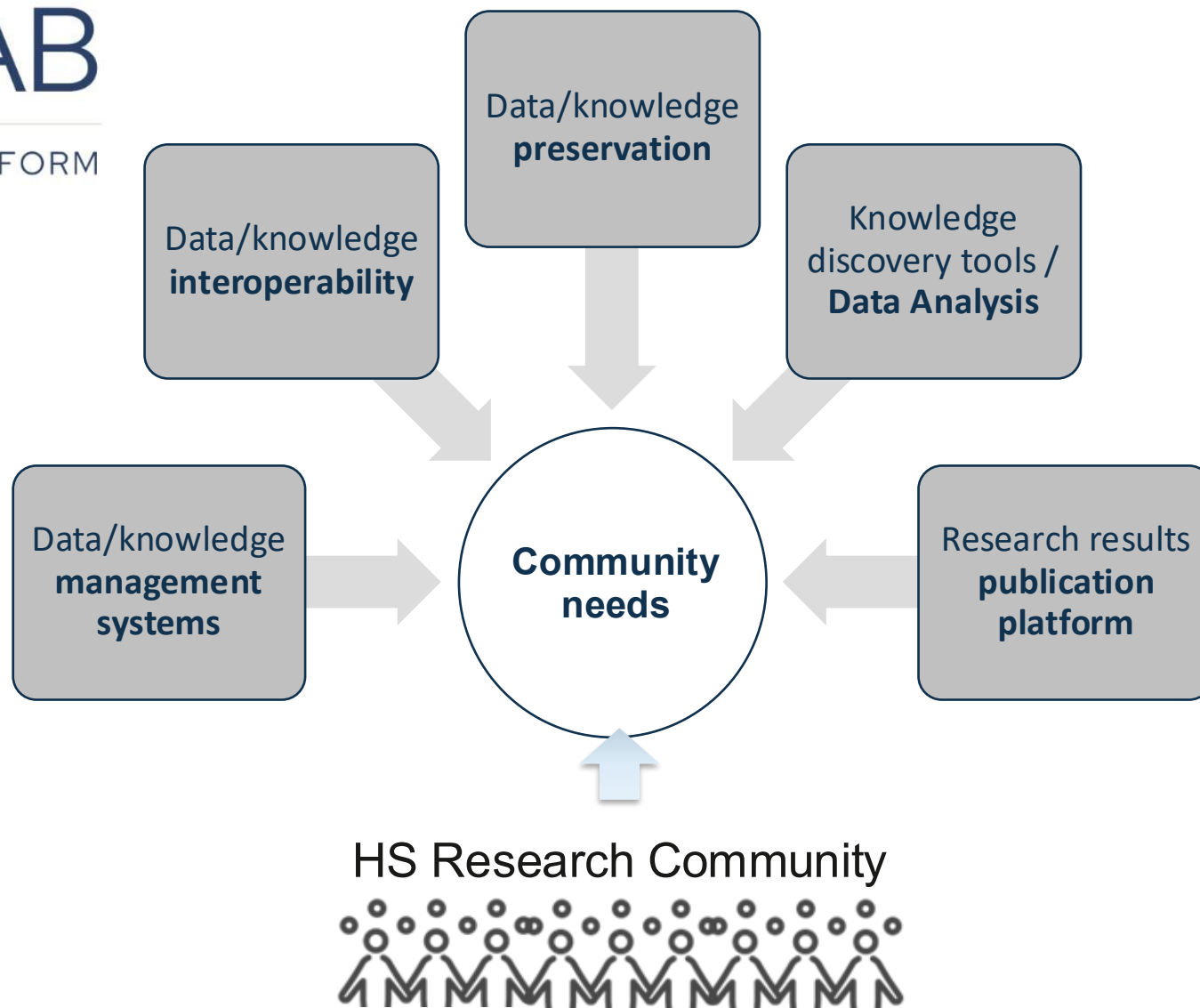
**FIXLAB**



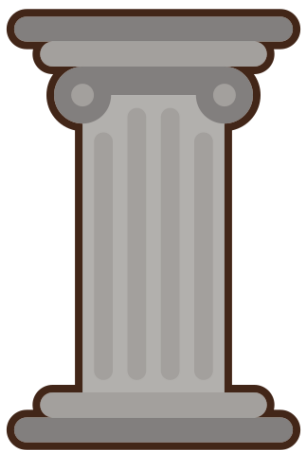
**MOLAB**



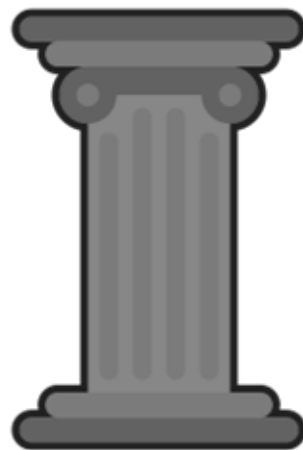
**DIGILAB**



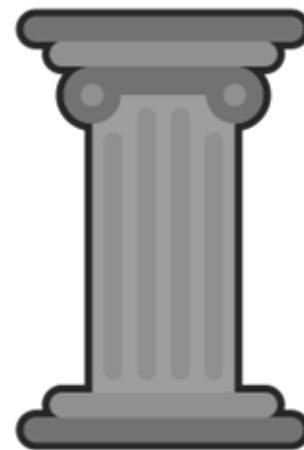
# DIGILAB.it **PILLARS**



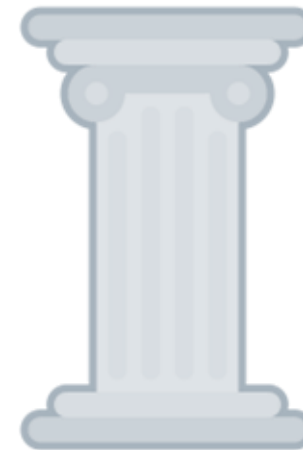
**Data management**



**Process management**



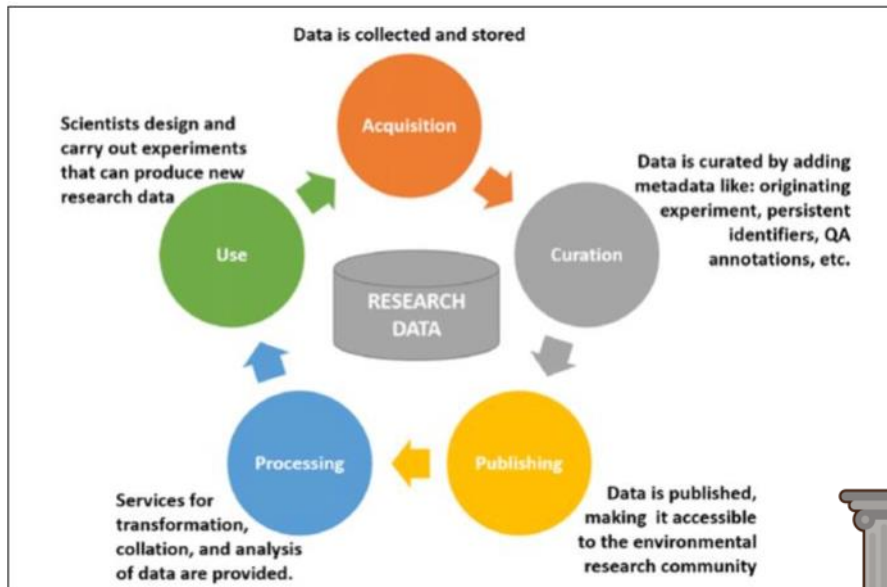
**Knowledge management**



**C-VRE**

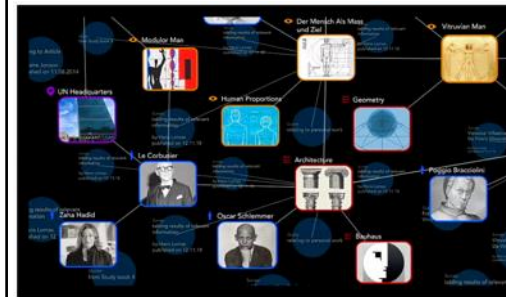
# DIGILAB.it: SEMANTIC HUB FOR HS KNOWLEDGE TO CAPTURE THE HS DATA DURING ITS WHOLE LIFECYCLE

**Data lifecycle management (DLM):** safeguarding data appropriately throughout its existence



SOURCE: **ENVRI REFERENCE MODEL:** (<https://envri.eu/wp-content/uploads/2020/07/The-ENVRI-Reference-Model.pdf>)

**Knowledge graphs** capture the different ways in which events, actors, and things connect and relate across time and space.



SOURCE: [Connect, communicate and represent knowledge with context](#)

Still image

**Research workflows** capture the scientific process in which the research take place



SOURCE: [Integrating research into the design process](#)

Dynamic image

# DIGILAB.it roadmap



**DATA PORTAL**

- Data Publishing
- Knowledge discovery

**DATA CATALOG**

- Data Management Platform

**C-VRE**

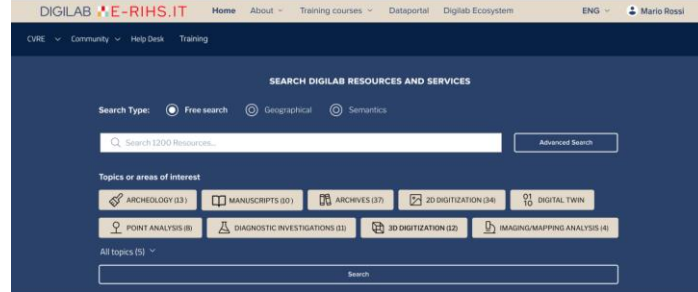
- Collaborative Virtual Research Environment

**COMMUNITY**

- Helpdesk
- Wiki/Forum
- e-Learning
- Survey

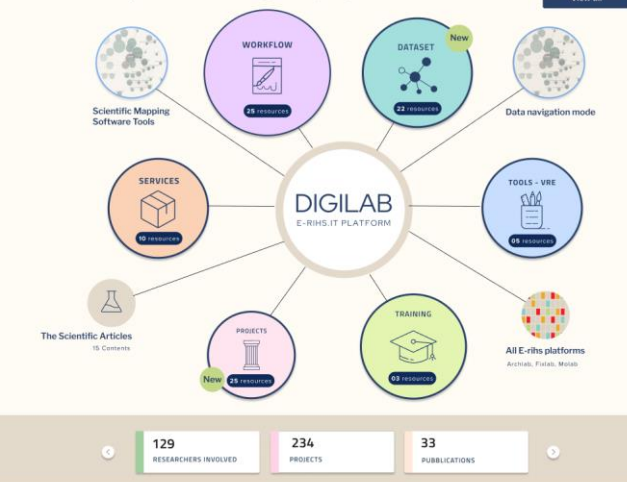


User-facing access and publishing layer

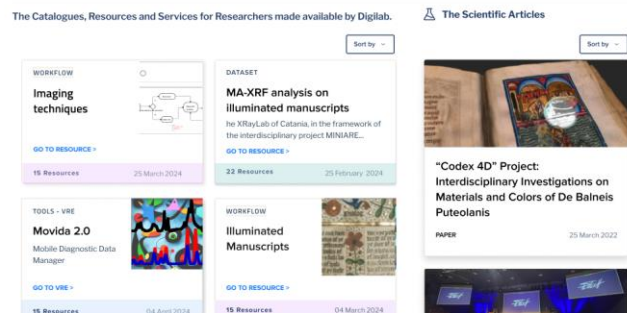


### Digilab Ecosystem

the Italian node of the European infrastructure for heritage science (E-RIHS)



### Resources and Services

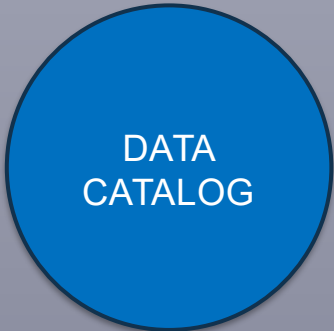


- Main entry point to the Digilab-IT ecosystem**
- Access to:** datasets, services, tools, and digital resources
- Multi-modal knowledge discovery:** free-text, geographic, and semantic search
- Publishing and dissemination** aligned with Open Science and FAIR principles
- Persistent identifier support** (e.g. DOI)

**DATA PORTAL**

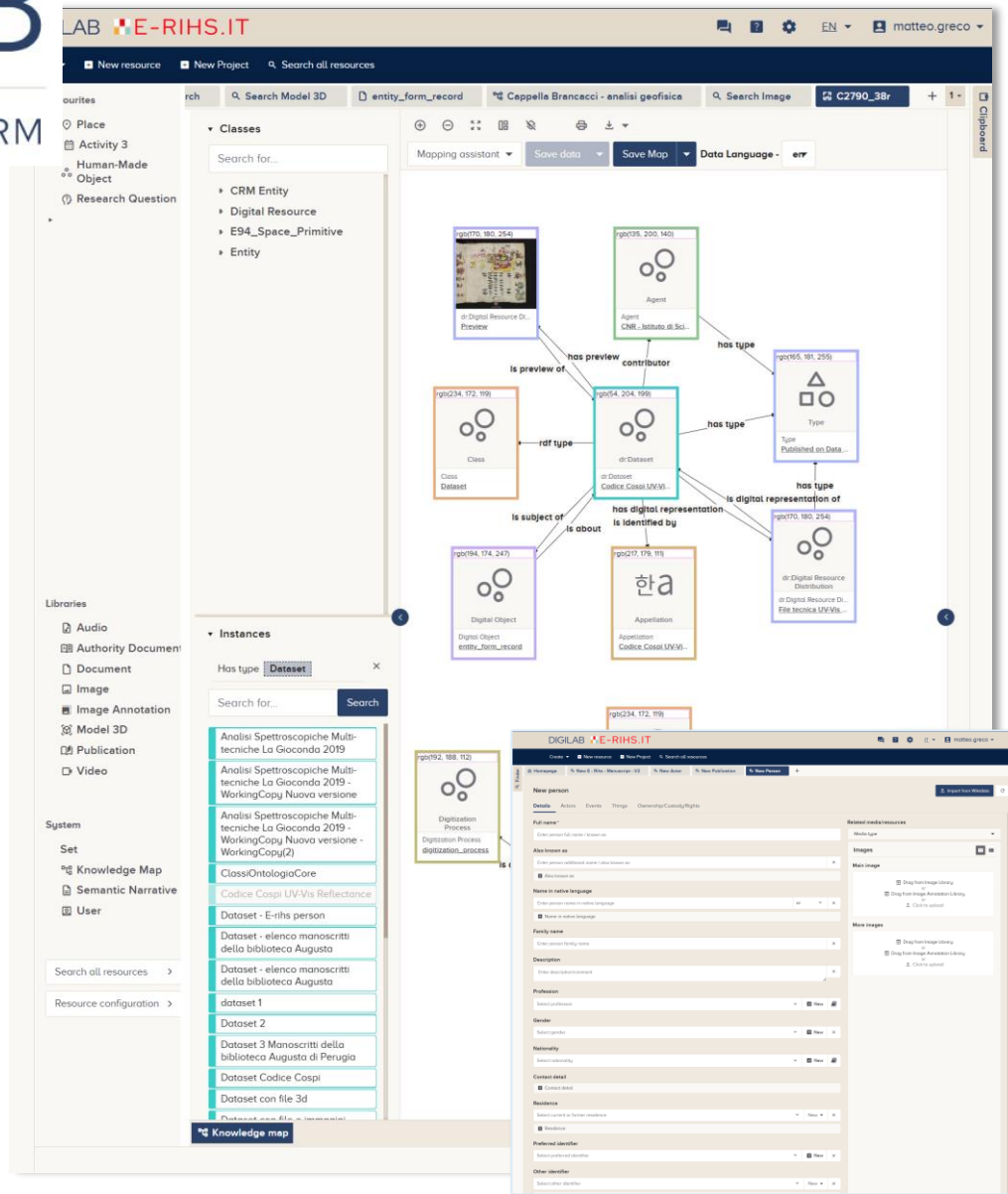
- Landing Platform
- Data Publishing
- Knowledge discovery

Semantic data management and knowledge graph backbone



**DATA CATALOG**

- Data/Knowledge Management Platform



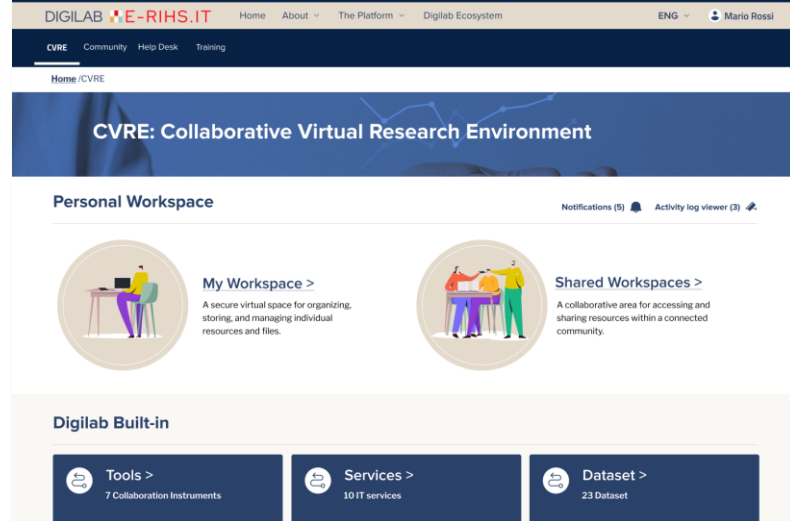
The screenshot displays the DIGILAB E-RIHS.IT platform interface. The top navigation bar includes 'LAB E-RIHS.IT', user information 'matteo.greco', and search options. The main area is divided into several sections:

- Classes:** A search bar and a list of classes including 'CRM Entity', 'Digital Resource', 'E94\_Space\_Primitive', and 'Entity'.
- Instances:** A search bar and a list of instances such as 'Analisi Spettroscopiche Multi-tecniche La Gioconda 2019', 'Codice Cospi UV-Vis Reflectance', and 'Dataset - E-rihs person'.
- Knowledge Map:** A central visualization of a knowledge graph with nodes and relationships. Nodes include 'Class Dataset', 'Digital Object', 'Agent', 'Type', and 'Appellation'. Relationships are labeled with terms like 'is preview of', 'has preview contributor', 'has type', 'rdf type', 'is subject of', 'is about', 'has digital representation', and 'is identified by'.
- Libraries:** A sidebar with categories like Audio, Authority Document, Document, Image, Image Annotation, Model 3D, Publication, and Video.
- System:** A sidebar with categories like Set, Knowledge Map, Semantic Narrative, and User.
- Search and Configuration:** Search bars for 'Search all resources' and 'Resource configuration'.

- Data and metadata management environment
- Knowledge graph life cycle management
- Ontology versioning
- No-code configuration of semantic Resource Models
- Integration with APIs, mappings, and ingestion pipelines



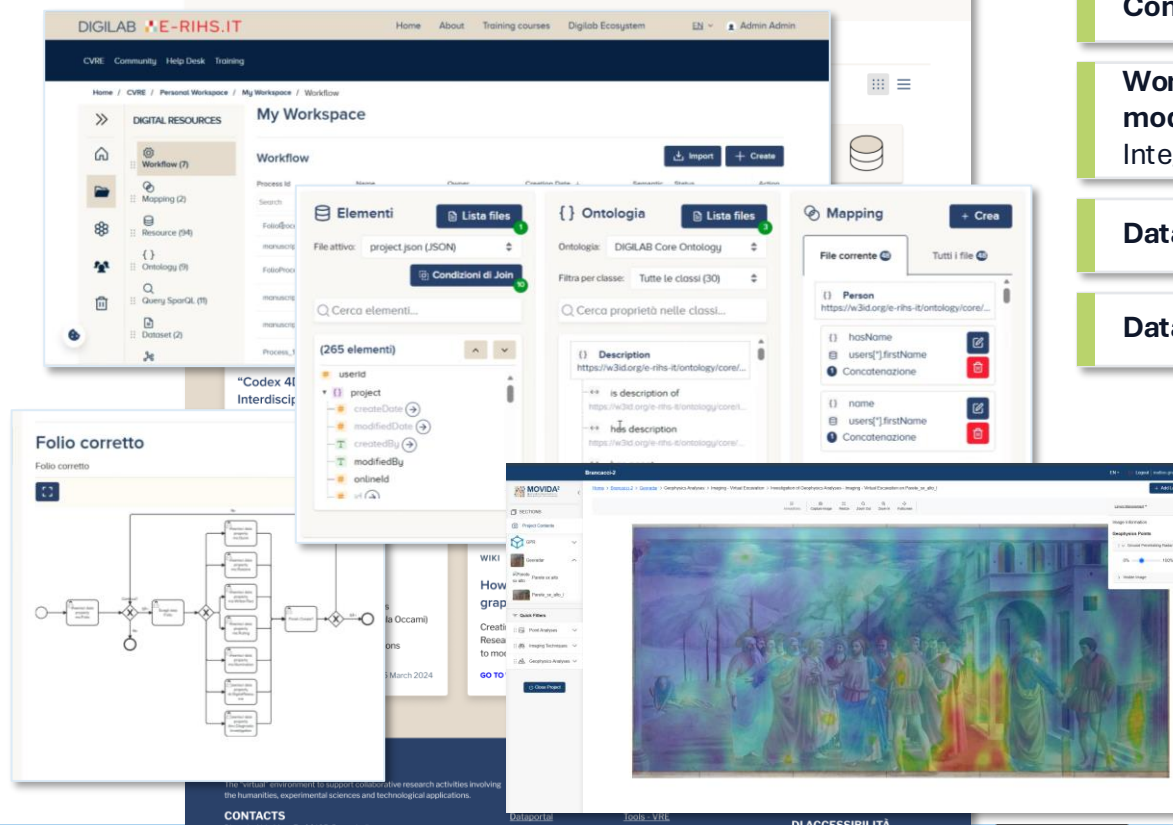
# Collaborative Virtual Research Environment




- Researcher dashboard
- Personal and shared workspaces
- Controlled access to tools and services
- Workflow orchestration and guided procedures modeled in BPMN
- Data analysis environments and analytical tools
- Data transformation and mapping pipelines

## C-VRE

- Collaborative Virtual Research Environment
- Workspaces
- Workflows
- Data Analysis
- Tools
- Services

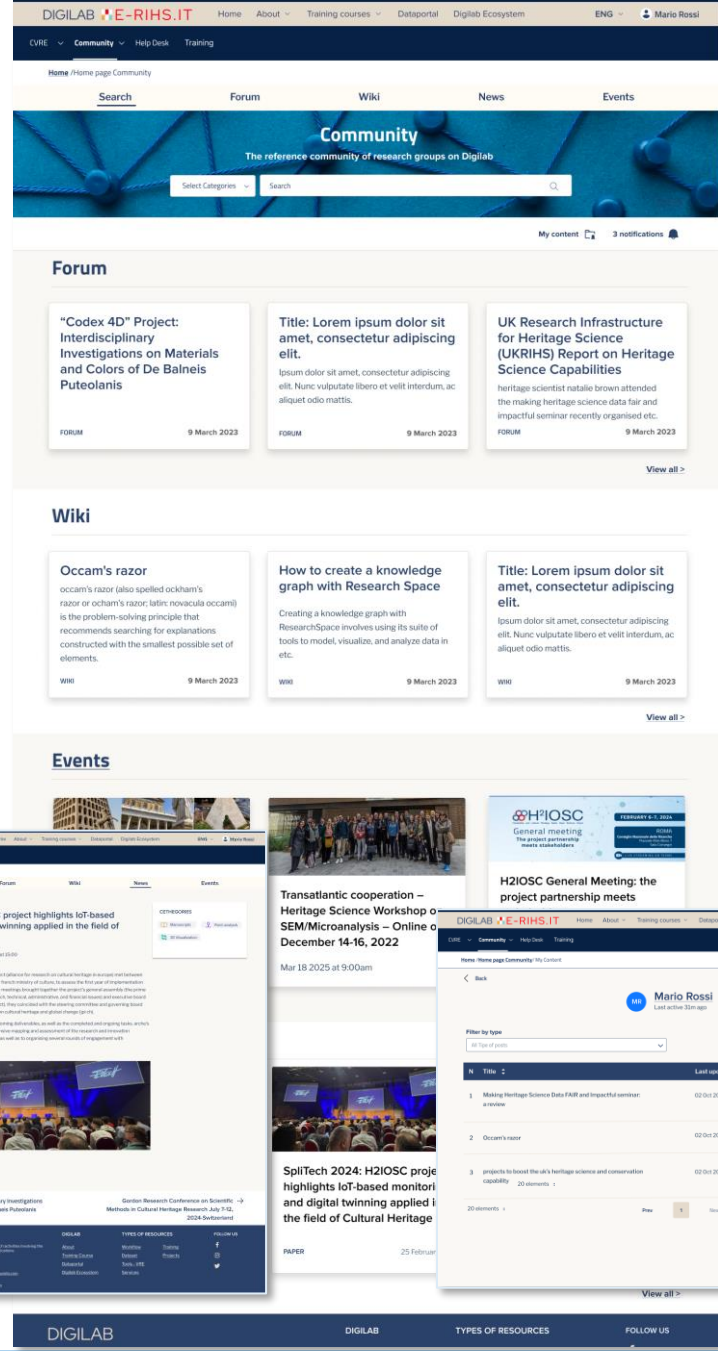


Community engagement, support, and knowledge-sharing layer



**COMMUNITY**

- Helpdesk
- Wiki/Forum
- e-Learning
- Survey



The screenshot shows the main interface of the DIGILAB E-RIHS.IT platform. It features a navigation bar with 'Home', 'Community', 'Help Desk', and 'Training'. Below the navigation, there are sections for 'Forum', 'Wiki', and 'Events'. The Forum section displays three articles, including one about the 'Codex 4D' Project. The Wiki section includes an article on 'Occam's razor' and a guide on 'How to create a knowledge graph with Research Space'. The Events section features a banner for 'Transatlantic cooperation – Heritage Science Workshop on SEM/Microanalysis – Online December 14-16, 2022' and another for 'H2IOSC General Meeting: the project partnership meets'. At the bottom, there is a 'DIGILAB' logo and a 'TYPES OF RESOURCES' section.

**Shared knowledge base, wiki-based**  
Documentation, knowledge exchange, and best-practice sharing

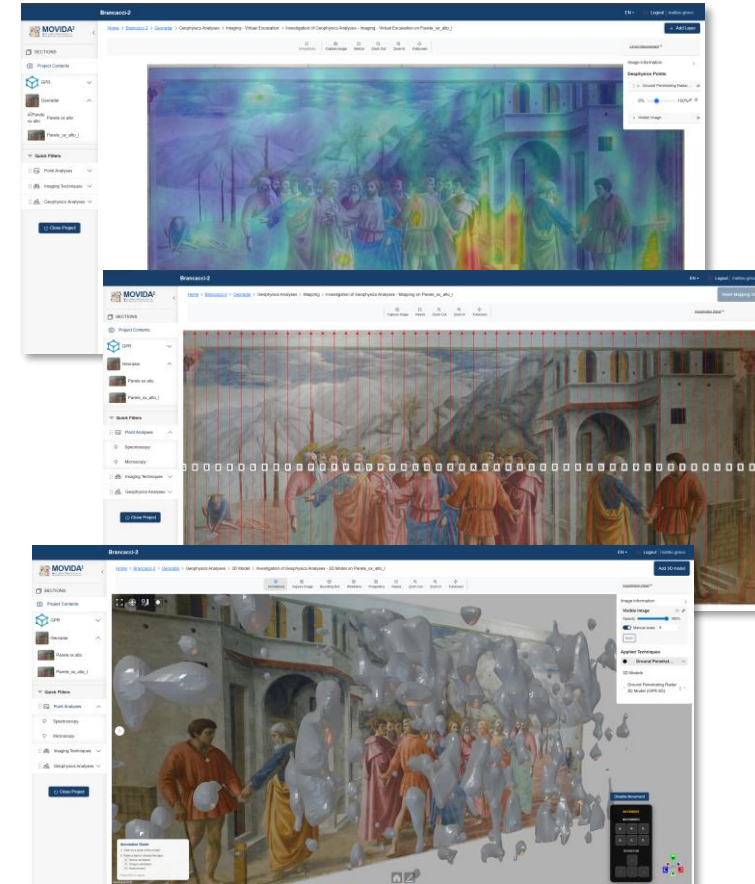
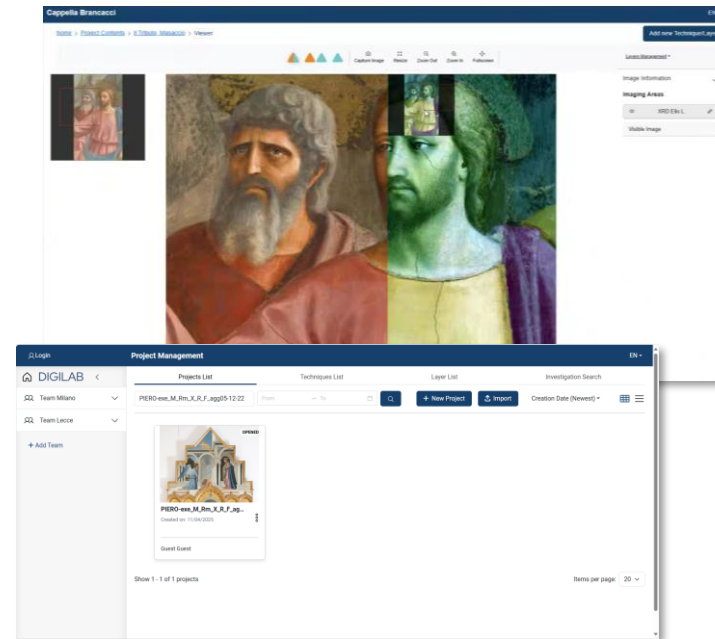
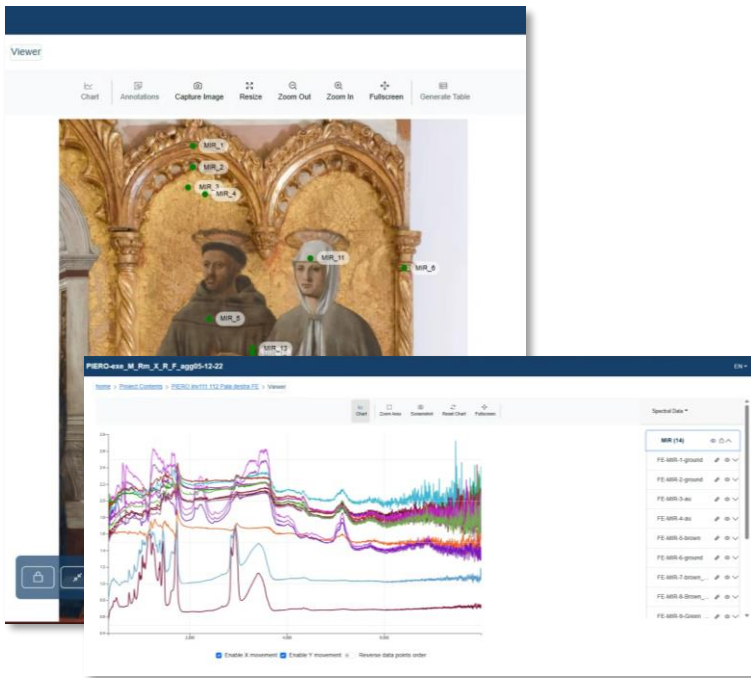
**eLearning environment**  
Training, onboarding and guidance materials

**User support through a ticketing system**

**News, Events and Forums**

**Survey system**

Support for: **Point-based analyses, Imaging/mapping techniques, Geophysics analyses**



# Interoperability in Digilab-IT: from technical integration to federation readiness

## Technical:

(the compatibility of access, exchange, and execution mechanisms across services.)

- Internal and federated authentication and authorization
- OpenAPI-based service integration
- Open, documented, machine-actionable data and metadata formats
- Persistent and resolvable identifiers
- Linked Data / Semantic Web-ready graph access and exchange mechanisms
- Metadata harvesting and synchronization services

## Semantic:

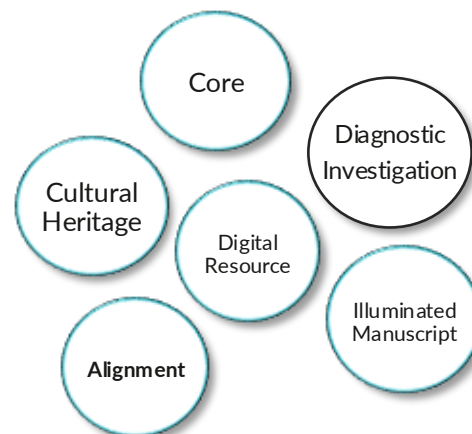
(the preservation of shared meaning across heterogeneous data, tools, and domains)

- Knowledge-graph-centric data organization
- Modular ontology network
- **Alignment layer with reference ontologies and shared conceptual models**
- Controlled vocabularies and semantic resource models
- Ontology versioning and semantic validation

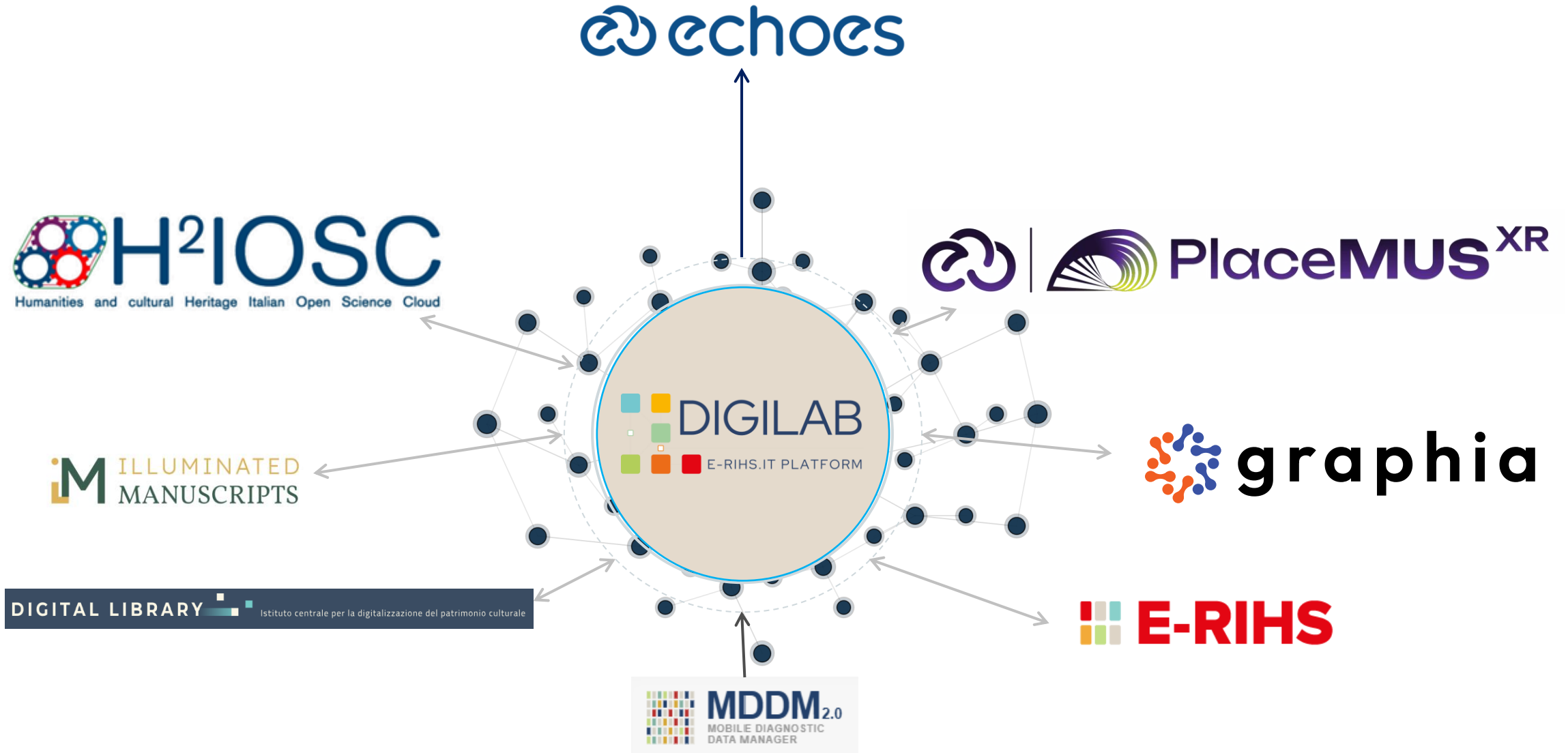
## Organizational:

(the alignment of workflows, roles, and service interactions required for real research operations)

- Shared workspaces and governed access
- **Reuse of common APIs, AAI, and metadata models by services and pilots**
- BPMN-based workflow orchestration
- Support, onboarding, and documentation for sustainable platform use

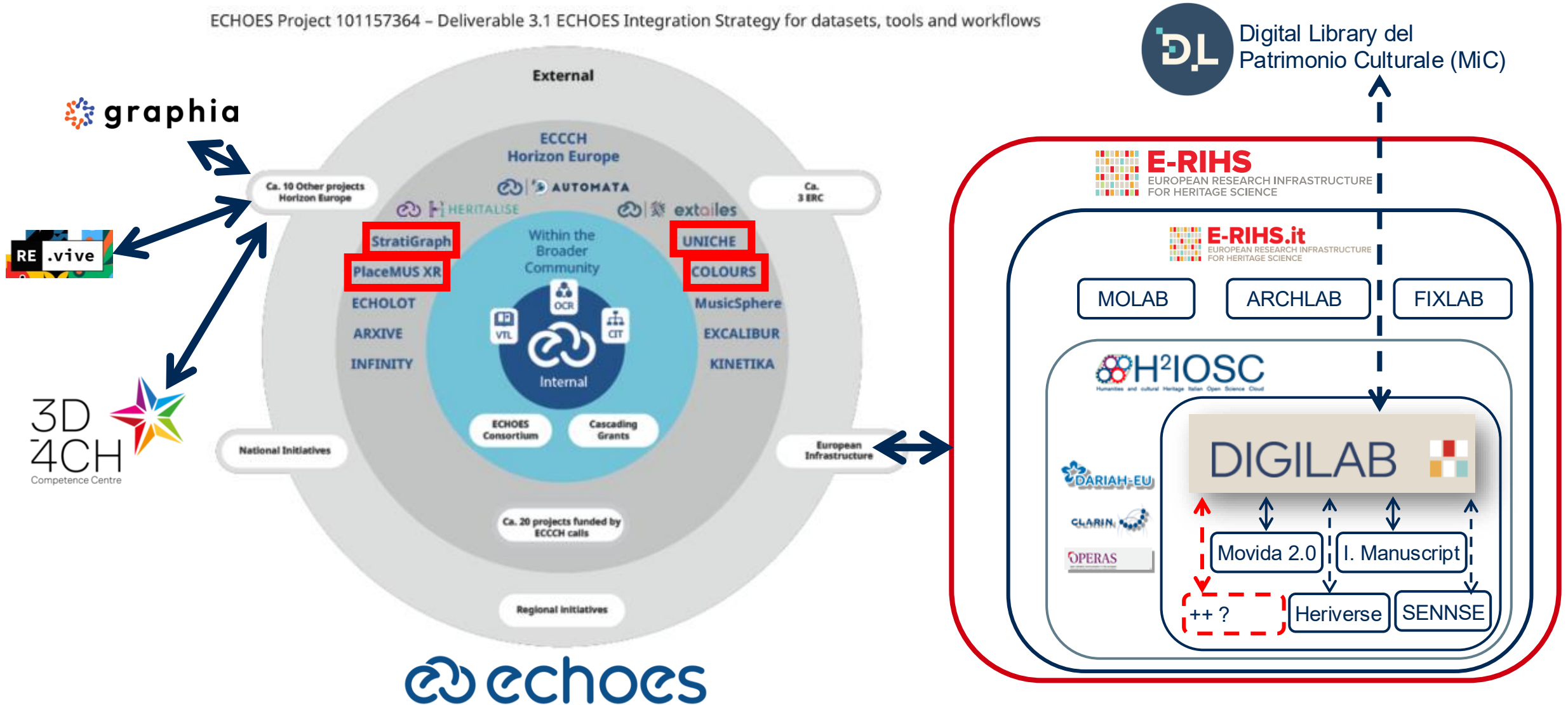


# Projects and Interoperability/Federation scenarios (current/possible)



# European Collaborative Cloud for Cultural Heritage (ECCCH)

ECHOES Project 101157364 – Deliverable 3.1 ECHOES Integration Strategy for datasets, tools and workflows



echoes

# DIGILAB@E-RIHS.it is not the answer.

*It is the infrastructure that makes answers possible.*

**Alberto Bucciero – CNR ISPC / Digital Heritage Lab**

[alberto.bucciero@cnr.it](mailto:alberto.bucciero@cnr.it)

[www.ispc.cnr.it](http://www.ispc.cnr.it) · [www.e-rihs.it](http://www.e-rihs.it) · [www.e-rihs.eu](http://www.e-rihs.eu)